

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

df=pd.read_excel('World_Bank_Projects_downloaded_6_15_2022.xls')
df.head()
```

World Bank Projects, data as of 06/14/2022 22:00:02 EST \

```
0          Project ID
1          id
2          P252350
3          P252338
4          P252337
```

```

      Unnamed: 1          Unnamed: 2 \
0          Region          Country
1      regionname      countryname
2  Eastern and Southern Africa  Republic of Rwanda
3      East Asia and Pacific  Democratic Republic of Timor-Leste
4  Eastern and Southern Africa      Republic of Kenya
```

```

      Unnamed: 3          Unnamed: 4 \
0  Project Status  Last Stage Reached Name
1  projectstatusdisplay  last_stage_reached_name
2          Active          Bank Approved
3          Active          Bank Approved
4          Dropped          Concept Review
```

```

      Unnamed: 5 \
0          Project Name
1          project_name
2      Rwanda Priority Skills for Growth (PSG)
3  Road Climate Resilience Project Second Additio...
4          Kenya Devolution and Service Delivery
```

```

      Unnamed: 6 \
0          Project Development Objective
1          pdo
2  The Program Development Objective is to expand...
3  The project will deliver sustainable climate r...
4  To improve citizen and government access to co...
```

```

      Unnamed: 7 \
0          Implementing Agency
1          impagency
2      Ministry Of Education
3  Ministry of Public Works, Transport, and Commu...
4          Kenya National Bureau of Statistics
```

Unnamed: 8 \

| | |
|---|------------------------------|
| 0 | Consultant Services Required |
| 1 | cons_serv_reqd_ind |
| 2 | NaN |
| 3 | NaN |
| 4 | NaN |

| | | | | |
|----------|---|-------------|-----|----------|
| | | Unnamed: 9 | ... | Unnamed: |
| 17 | \ | | | |
| 0 | | Project URL | ... | Grant |
| Amount | | | | |
| 1 | | url | ... | |
| grantamt | | | | |
| 2 | http://projects.worldbank.org/P252350/null?lan... | | ... | |
| 0 | | | | |
| 3 | http://projects.worldbank.org/P252338/null?lan... | | ... | |
| 0 | | | | |
| 4 | http://projects.worldbank.org/P252337/null?lan... | | ... | |
| 340000 | | | | |

| | | |
|--------------|---|---------------------|
| | | Unnamed: 18 |
| Unnamed: 19 | \ | |
| 0 | | Borrower |
| Instrument | | Lending |
| 1 | | borrower |
| lendinginstr | | |
| 2 | Ministry of Finance and Economic Planning | Program-for-Results |
| Financing | | |
| 3 | Ministry of Finance | Investment Project |
| Financing | | |
| 4 | Kenya National Bureau of Statistics | Investment Project |
| Financing | | |

| | | | | | |
|---|-----------------------------------|-------------------------------|--|-------------|---|
| | | Unnamed: 20 | | Unnamed: 21 | \ |
| 0 | Environmental Assessment Category | Environmental and Social Risk | | | |
| 1 | envassesmentcategorycode | esrc_ovrl_risk_rate | | | |
| 2 | NaN | NaN | | | |
| 3 | B | NaN | | | |
| 4 | C | NaN | | | |

| | | |
|----------------|-----------------------------------|-------------------------|
| | | Unnamed: 22 |
| Unnamed: 23 | \ | |
| 0 | | Sector 1 |
| Sector 2 | | |
| 1 | | sector1 |
| sector2 | | |
| 2 | Public Administration - Education | Tertiary |
| Education | | |
| 3 | Rural and Inter-Urban Roads | Public Administration - |
| Transportation | | |
| 4 | | NaN |

NaN

```
0      Unnamed: 24 \
1      Sector 3
2      sector3
2 Workforce Development and Vocational Education
3      NaN
4      NaN
```

```
      Unnamed: 25      Unnamed: 26
0      Theme 1      Theme 2
1      theme1      theme2
2      Education for the knowledge economy      NaN
3      NaN      NaN
4 Public expenditure, financial management and p... Decentralization
```

[5 rows x 27 columns]

```
df_n=df.rename(columns=df.iloc[0])
```

```
df_n=df_n.drop(index=0)
```

```
df_n=df_n.drop(index=1)
```

df_n

```
      Project ID      Region \
2      P252350 Eastern and Southern Africa
3      P252338      East Asia and Pacific
4      P252337 Eastern and Southern Africa
5      P179254 Western and Central Africa
6      P179238 Western and Central Africa
...      ...
21540      P000019      Africa
21541      P000017      Africa
21542      P000010      Africa
21543      P000003      Africa
21544      P000001      Africa
```

```
      Country Project Status \
2      Republic of Rwanda      Active
3      Democratic Republic of Timor-Leste      Active
4      Republic of Kenya      Dropped
5      Republic of Cameroon      Pipeline
6      Republic of Chad      Pipeline
```

| | | |
|-------|--------|---------|
| ... | ... | ... |
| 21540 | Africa | Dropped |
| 21541 | Africa | Closed |
| 21542 | Africa | Closed |
| 21543 | Africa | Closed |
| 21544 | Africa | Closed |

| | |
|---|---------------------------|
| | Last Stage Reached Name \ |
| 2 | Bank Approved |
| 3 | Bank Approved |
| 4 | Concept Review |
| 5 | NaN |
| 6 | NaN |

| | |
|-------|---------------|
| ... | ... |
| 21540 | NaN |
| 21541 | Bank Approved |
| 21542 | NaN |
| 21543 | Bank Approved |
| 21544 | Bank Approved |

| | |
|---|---|
| | Project Name \ |
| 2 | Rwanda Priority Skills for Growth (PSG) |
| 3 | Road Climate Resilience Project Second Additio... |
| 4 | Kenya Devolution and Service Delivery |
| 5 | Cameroon Health System Strengthening Project |
| 6 | Chad Agribusiness and Rural Transformation Pro... |

| | |
|-------|---|
| ... | ... |
| 21540 | UEMOA Regional Financial Sector Project |
| 21541 | Engineering and Technical Assistance Project |
| 21542 | Regional Development Project (03) |
| 21543 | REIMP(CEN.ENV.INFO) |
| 21544 | West Africa Pilot Community-based Natural Reso... |

| | |
|---|---|
| | Project Development Objective \ |
| 2 | The Program Development Objective is to expand... |
| 3 | The project will deliver sustainable climate r... |
| 4 | To improve citizen and government access to co... |
| 5 | To increase utilization and quality of reprodu... |
| 6 | The project Development Objective (PDO) is "to... |

| | |
|-------|-----|
| ... | ... |
| 21540 | NaN |
| 21541 | NaN |
| 21542 | NaN |
| 21543 | NaN |
| 21544 | NaN |

| | |
|---|---|
| | Implementing Agency \ |
| 2 | Ministry Of Education |
| 3 | Ministry of Public Works, Transport, and Commu... |
| 4 | Kenya National Bureau of Statistics |

| | |
|-------|-----|
| 5 | NaN |
| 6 | NaN |
| ... | ... |
| 21540 | NaN |
| 21541 | NaN |
| 21542 | NaN |
| 21543 | NaN |
| 21544 | NaN |

| Consultant Services Required \ | |
|--------------------------------|-----|
| 2 | NaN |
| 3 | NaN |
| 4 | NaN |
| 5 | NaN |
| 6 | NaN |
| ... | ... |
| 21540 | NaN |
| 21541 | NaN |
| 21542 | NaN |
| 21543 | NaN |
| 21544 | NaN |

| Amount \ | Project URL | ... | Grant |
|----------|---|-----|-------|
| 2 | http://projects.worldbank.org/P252350/null?lan... | ... | ... |
| 0 | | | |
| 3 | http://projects.worldbank.org/P252338/null?lan... | ... | ... |
| 0 | | | |
| 4 | http://projects.worldbank.org/P252337/null?lan... | ... | ... |
| 340000 | | | |
| 5 | http://projects.worldbank.org/P179254/null?lan... | ... | ... |
| 0 | | | |
| 6 | http://projects.worldbank.org/P179238/null?lan... | ... | ... |
| 0 | | | |
| ... | | ... | ... |
| ... | | | |
| 21540 | http://projects.worldbank.org/P000019/null?lan... | ... | ... |
| 0 | | | |
| 21541 | http://projects.worldbank.org/P000017/engineer... | ... | ... |
| 0 | | | |
| 21542 | http://projects.worldbank.org/P000010/null?lan... | ... | ... |
| 0 | | | |
| 21543 | http://projects.worldbank.org/P000003/reimpce... | ... | ... |
| 4100000 | | | |
| 21544 | http://projects.worldbank.org/P000001/west-afr... | ... | ... |
| 4400000 | | | |

| Borrower \ | |
|------------|---|
| 2 | Ministry of Finance and Economic Planning |
| 3 | Ministry of Finance |

| | | |
|-------|-------------------------------------|-----|
| 4 | Kenya National Bureau of Statistics | |
| 5 | | NaN |
| 6 | | NaN |
| ... | | ... |
| 21540 | | NaN |
| 21541 | | NaN |
| 21542 | | NaN |
| 21543 | | NaN |
| 21544 | | NaN |

| Lending Instrument Environmental Assessment Category | | |
|--|-------------------------------|-----|
| \ | | |
| 2 | Program-for-Results Financing | NaN |
| 3 | Investment Project Financing | B |
| 4 | Investment Project Financing | C |
| 5 | Investment Project Financing | NaN |
| 6 | Investment Project Financing | NaN |
| ... | ... | ... |
| 21540 | Specific Investment Loan | C |
| 21541 | Technical Assistance Loan | C |
| 21542 | Financial Intermediary Loan | B |
| 21543 | Specific Investment Loan | C |
| 21544 | Specific Investment Loan | B |

| Environmental and Social Risk | | Sector 1 |
|-------------------------------|-----|-----------------------------------|
| \ | | |
| 2 | NaN | Public Administration - Education |
| 3 | NaN | Rural and Inter-Urban Roads |
| 4 | NaN | NaN |
| 5 | NaN | NaN |
| 6 | NaN | NaN |
| ... | ... | ... |

| | | |
|-------|-----|-----------------------------|
| 21540 | NaN | NaN |
| 21541 | NaN | Power |
| 21542 | NaN | Banking Institutions |
| 21543 | NaN | Other Public Administration |
| 21544 | NaN | Sub-National Government |

| | | |
|-------|---|--------------------|
| | | Sector 2 \ |
| 2 | | Tertiary Education |
| 3 | Public Administration - Transportation | |
| 4 | | NaN |
| 5 | | NaN |
| 6 | | NaN |
| ... | | ... |
| 21540 | | NaN |
| 21541 | | NaN |
| 21542 | Other Agriculture, Fishing and Forestry | |
| 21543 | Other Information and Communications Technologies | |
| 21544 | Social Protection | |

| | | |
|-------|--|------------|
| | | Sector 3 \ |
| 2 | Workforce Development and Vocational Education | |
| 3 | | NaN |
| 4 | | NaN |
| 5 | | NaN |
| 6 | | NaN |
| ... | | ... |
| 21540 | | NaN |
| 21541 | | NaN |
| 21542 | Other Transportation | |
| 21543 | Social Protection | |
| 21544 | Other Agriculture, Fishing and Forestry | |

| | | |
|-------|---|-----------|
| | | Theme 1 \ |
| 2 | Education for the knowledge economy | |
| 3 | | NaN |
| 4 | Public expenditure, financial management and p... | |
| 5 | | NaN |
| 6 | | NaN |
| ... | | ... |
| 21540 | | NaN |
| 21541 | Legal institutions for a market economy | |
| 21542 | Environmental policies and institutions | |
| 21543 | Land administration and management | |

21544 Rural services and infrastructure

```

Theme 2
2      NaN
3      NaN
4      Decentralization
5      NaN
6      NaN
...
21540      NaN
21541      Regional integration
21542 Other financial and private sector development
21543      Environmental policies and institutions
21544      Biodiversity
```

[21543 rows x 27 columns]

df_n.shape

(21543, 27)

df_n.info()

<class 'pandas.core.frame.DataFrame'>

Int64Index: 21543 entries, 2 to 21544

Data columns (total 27 columns):

| # | Column | Non-Null Count | Dtype |
|----|-----------------------------------|----------------|--------|
| 0 | Project ID | 21543 non-null | object |
| 1 | Region | 21540 non-null | object |
| 2 | Country | 21540 non-null | object |
| 3 | Project Status | 21538 non-null | object |
| 4 | Last Stage Reached Name | 15197 non-null | object |
| 5 | Project Name | 21541 non-null | object |
| 6 | Project Development Objective | 11579 non-null | object |
| 7 | Implementing Agency | 6650 non-null | object |
| 8 | Consultant Services Required | 1483 non-null | object |
| 9 | Project URL | 21541 non-null | object |
| 10 | Board Approval Date | 18785 non-null | object |
| 11 | Project Closing Date | 16585 non-null | object |
| 12 | Financing Type | 13037 non-null | object |
| 13 | Current Project Cost | 21394 non-null | object |
| 14 | IBRD Commitment | 21541 non-null | object |
| 15 | IDA Commitment | 21541 non-null | object |
| 16 | Total IDA and IBRD Commitment | 21541 non-null | object |
| 17 | Grant Amount | 21541 non-null | object |
| 18 | Borrower | 6413 non-null | object |
| 19 | Lending Instrument | 21293 non-null | object |
| 20 | Environmental Assessment Category | 13082 non-null | object |
| 21 | Environmental and Social Risk | 1275 non-null | object |
| 22 | Sector 1 | 19627 non-null | object |


```

23 Sector 2          11176 non-null object
24 Sector 3          7565 non-null object
25 Theme 1          12255 non-null object
26 Theme 2          9895 non-null object
dtypes: object(27)
memory usage: 4.6+ MB

```

```
df_n.describe().T
```

| | count | unique | \ |
|-----------------------------------|-------|--------|---|
| Project ID | 21543 | 21543 | |
| Region | 21540 | 11 | |
| Country | 21540 | 207 | |
| Project Status | 21538 | 4 | |
| Last Stage Reached Name | 15197 | 5 | |
| Project Name | 21541 | 18605 | |
| Project Development Objective | 11579 | 9861 | |
| Implementing Agency | 6650 | 4625 | |
| Consultant Services Required | 1483 | 3 | |
| Project URL | 21541 | 21541 | |
| Board Approval Date | 18785 | 5984 | |
| Project Closing Date | 16585 | 1831 | |
| Financing Type | 13037 | 4 | |
| Current Project Cost | 21394 | 4583 | |
| IBRD Commitment | 21541 | 1301 | |
| IDA Commitment | 21541 | 1237 | |
| Total IDA and IBRD Commitment | 21541 | 1912 | |
| Grant Amount | 21541 | 969 | |
| Borrower | 6413 | 2187 | |
| Lending Instrument | 21293 | 18 | |
| Environmental Assessment Category | 13082 | 8 | |
| Environmental and Social Risk | 1275 | 4 | |
| Sector 1 | 19627 | 159 | |
| Sector 2 | 11176 | 102 | |
| Sector 3 | 7565 | 90 | |
| Theme 1 | 12255 | 73 | |
| Theme 2 | 9895 | 73 | |

```

top \
Project ID
P252350
Region                                Latin America
and Caribbean
Country
Republic of India
Project Status
Closed
Last Stage Reached Name
Bank Approved

```

| | | |
|---|-------------------------------|--------------------|
| Project Name | | |
| Education Project (02) | | |
| Project Development Objective | Project Development Objective | |
| (Note: will be d... | | |
| Implementing Agency | | |
| Ministry of Finance | | |
| Consultant Services Required | | |
| Y | | |
| Project URL | | |
| http://projects.worldbank.org/P252350/null?lan... | | |
| Board Approval Date | | 2020- |
| 04-02T00:00:00Z | | |
| Project Closing Date | | |
| 12/31/2022 12:00:00 AM | | |
| Financing Type | | |
| Other | | |
| Current Project Cost | | |
| 100000000 | | |
| IBRD Commitment | | |
| 0 | | |
| IDA Commitment | | |
| 0 | | |
| Total IDA and IBRD Commitment | | |
| 0 | | |
| Grant Amount | | |
| 0 | | |
| Borrower | | |
| Ministry of Finance | | |
| Lending Instrument | | Specific |
| Investment Loan | | |
| Environmental Assessment Category | | |
| B | | |
| Environmental and Social Risk | | |
| Substantial | | |
| Sector 1 | | Central Government |
| (Central Agencies) | | |
| Sector 2 | | Central Government |
| (Central Agencies) | | |
| Sector 3 | | Central Government |
| (Central Agencies) | | |
| Theme 1 | | Rural services and |
| infrastructure | | |
| Theme 2 | | Rural services and |
| infrastructure | | |
| | freq | |
| Project ID | 1 | |
| Region | 3766 | |
| Country | 918 | |
| Project Status | 15783 | |

| | |
|-----------------------------------|-------|
| Last Stage Reached Name | 13495 |
| Project Name | 70 |
| Project Development Objective | 47 |
| Implementing Agency | 227 |
| Consultant Services Required | 912 |
| Project URL | 1 |
| Board Approval Date | 28 |
| Project Closing Date | 160 |
| Financing Type | 4835 |
| Current Project Cost | 531 |
| IBRD Commitment | 13331 |
| IDA Commitment | 12013 |
| Total IDA and IBRD Commitment | 4264 |
| Grant Amount | 17626 |
| Borrower | 653 |
| Lending Instrument | 8021 |
| Environmental Assessment Category | 5862 |
| Environmental and Social Risk | 542 |
| Sector 1 | 1815 |
| Sector 2 | 1632 |
| Sector 3 | 760 |
| Theme 1 | 914 |
| Theme 2 | 713 |

handling missing values and feature engineering

```
df_n.isnull().sum()
```

| | |
|-----------------------------------|-------|
| Project ID | 0 |
| Region | 3 |
| Country | 3 |
| Project Status | 5 |
| Last Stage Reached Name | 6346 |
| Project Name | 2 |
| Project Development Objective | 9964 |
| Implementing Agency | 14893 |
| Consultant Services Required | 20060 |
| Project URL | 2 |
| Board Approval Date | 2758 |
| Project Closing Date | 4958 |
| Financing Type | 8506 |
| Current Project Cost | 149 |
| IBRD Commitment | 2 |
| IDA Commitment | 2 |
| Total IDA and IBRD Commitment | 2 |
| Grant Amount | 2 |
| Borrower | 15130 |
| Lending Instrument | 250 |
| Environmental Assessment Category | 8461 |
| Environmental and Social Risk | 20268 |
| Sector 1 | 1916 |

```
Sector 2          10367
Sector 3          13978
Theme 1           9288
Theme 2          11648
dtype: int64
```

here i don't need project id becuae it didn't effect data, and also mention two columns for place so we remove region also. And I delete URL.

```
df_new=df_n.drop(df_n[['Project ID','Region','Project URL']],axis=1)
```

```
df_new.head()
```

| | Reached Name \ | Country | Project Status | Last Stage |
|---|----------------|------------------------------------|----------------|------------|
| 2 | Approved | Republic of Rwanda | Active | Bank |
| 3 | Approved | Democratic Republic of Timor-Leste | Active | Bank |
| 4 | Review | Republic of Kenya | Dropped | Concept |
| 5 | NaN | Republic of Cameroon | Pipeline | |
| 6 | NaN | Republic of Chad | Pipeline | |

| | | Project Name \ |
|---|------|--|
| 2 | | Rwanda Priority Skills for Growth (PSG) |
| 3 | Road | Climate Resilience Project Second Additio... |
| 4 | | Kenya Devolution and Service Delivery |
| 5 | | Cameroon Health System Strengthening Project |
| 6 | Chad | Agribusiness and Rural Transformation Pro... |

| | | Project Development Objective \ |
|---|--|---|
| 2 | | The Program Development Objective is to expand... |
| 3 | | The project will deliver sustainable climate r... |
| 4 | | To improve citizen and government access to co... |
| 5 | | To increase utilization and quality of reprodu... |
| 6 | | The project Development Objective (PDO) is "to... |

| | | Implementing Agency \ |
|---|--|---|
| 2 | | Ministry Of Education |
| 3 | | Ministry of Public Works, Transport, and Commu... |
| 4 | | Kenya National Bureau of Statistics |
| 5 | | NaN |
| 6 | | NaN |

| | Consultant Services Required | Board Approval Date | Project Closing |
|---|------------------------------|----------------------|-----------------|
| 2 | NaN | 2017-07-05T00:00:00Z | 9/30/2023 |

12:00:00 AM

| | | |
|-----|-----|----------------------|
| 3 | NaN | 2017-04-27T00:00:00Z |
| NaN | | |
| 4 | NaN | NaN |
| NaN | | |
| 5 | NaN | NaN |
| NaN | | |
| 6 | NaN | NaN |
| NaN | | |

| Financing Type | ... | Grant Amount | |
|----------------|-----|--------------|---|
| Borrower \ | | | |
| 2 | IDA | ... | 0 Ministry of Finance and Economic Planning |
| 3 | NaN | ... | 0 Ministry of Finance |
| 4 | NaN | ... | 340000 Kenya National Bureau of Statistics |
| 5 | NaN | ... | 0 |
| NaN | | | |
| 6 | NaN | ... | 0 |
| NaN | | | |

| Lending Instrument | Environmental Assessment Category | \ |
|---------------------------------|-----------------------------------|---|
| 2 Program-for-Results Financing | NaN | |
| 3 Investment Project Financing | B | |
| 4 Investment Project Financing | C | |
| 5 Investment Project Financing | NaN | |
| 6 Investment Project Financing | NaN | |

| Environmental and Social Risk | Sector 1 | \ |
|-------------------------------|----------|-----------------------------------|
| 2 | NaN | Public Administration - Education |
| 3 | NaN | Rural and Inter-Urban Roads |
| 4 | NaN | NaN |
| 5 | NaN | NaN |
| 6 | NaN | NaN |

| Sector 2 | \ |
|----------|--|
| 2 | Tertiary Education |
| 3 | Public Administration - Transportation |
| 4 | NaN |
| 5 | NaN |
| 6 | NaN |

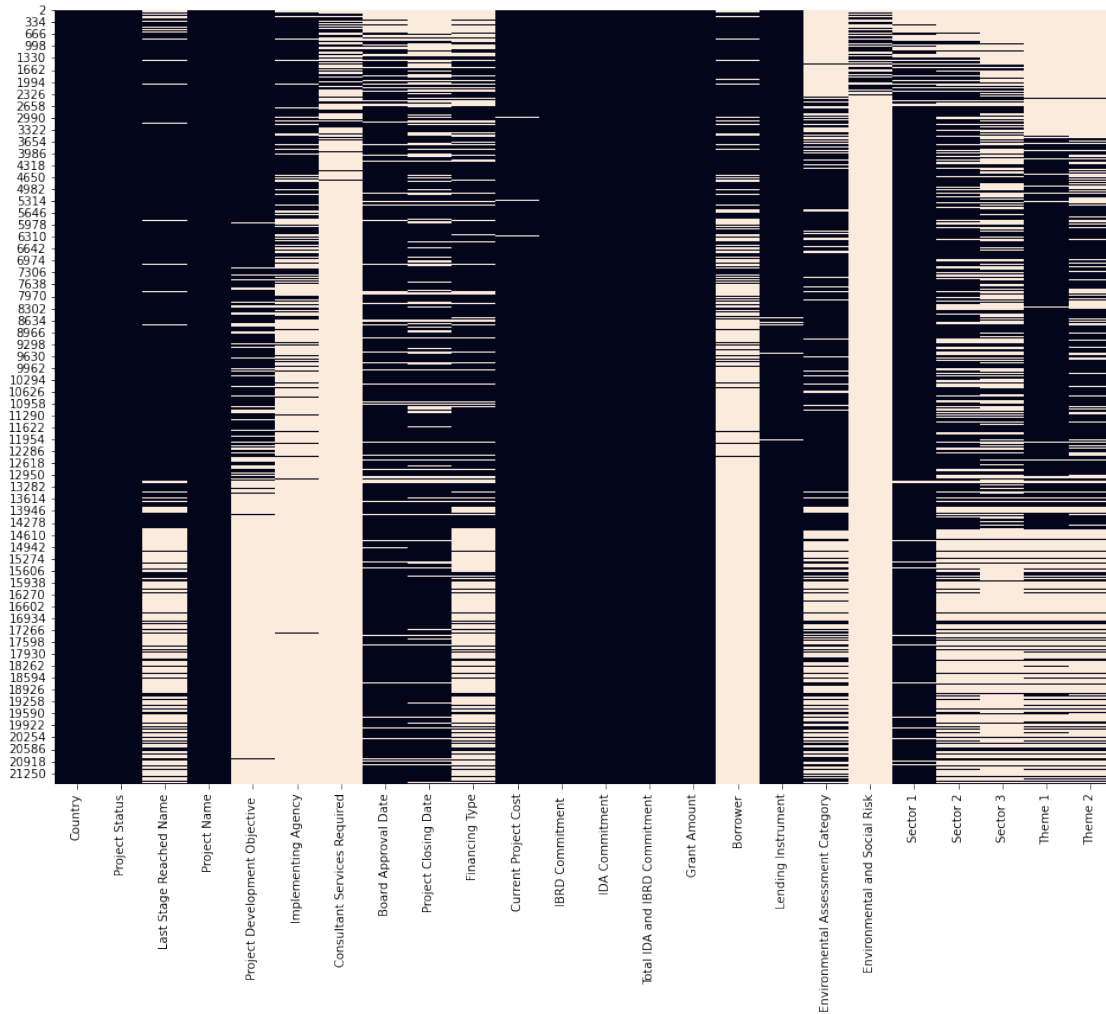
| Sector 3 | \ |
|----------|--|
| 2 | Workforce Development and Vocational Education |
| 3 | NaN |
| 4 | NaN |
| 5 | NaN |
| 6 | NaN |

| | Theme 1 | Theme 2 |
|---|---|------------------|
| 2 | Education for the knowledge economy | NaN |
| 3 | NaN | NaN |
| 4 | Public expenditure, financial management and p... | Decentralization |
| 5 | NaN | NaN |
| 6 | NaN | NaN |

[5 rows x 24 columns]

```
plt.figure(figsize=(16,12))
sns.heatmap(df_new.isnull(), cbar=False)
```

<AxesSubplot:>



In this dataset i got data 21543 rows,so in this is there any data which have more than half of null that column i will remove otherwise it will effect on my further process

```
df_new.isna().sum()
```

| | |
|-------------------------------|-------|
| Country | 3 |
| Project Status | 5 |
| Last Stage Reached Name | 6346 |
| Project Name | 2 |
| Project Development Objective | 9964 |
| Implementing Agency | 14893 |
| Consultant Services Required | 20060 |
| Board Approval Date | 2758 |
| Project Closing Date | 4958 |
| Financing Type | 8506 |
| Current Project Cost | 149 |
| IBRD Commitment | 2 |
| IDA Commitment | 2 |
| Total IDA and IBRD Commitment | 2 |

```

Grant Amount                2
Borrower                    15130
Lending Instrument          250
Environmental Assessment Category 8461
Environmental and Social Risk 20268
Sector 1                    1916
Sector 2                    10367
Sector 3                    13978
Theme 1                     9288
Theme 2                    11648
dtype: int64

```

so I take Implementing Agency, Consultant Services Required, Borrower, Environmental and Social Risk, Sector 3 & Theme 2 for remove

```

df_new2=df_new.drop(df_new[['Implementing Agency','Consultant Services
Required','Borrower','Environmental and Social Risk','Sector 3','Theme
2']],axis=1)

```

df_new2

| | Country | Project Status \ |
|-------|------------------------------------|------------------|
| 2 | Republic of Rwanda | Active |
| 3 | Democratic Republic of Timor-Leste | Active |
| 4 | Republic of Kenya | Dropped |
| 5 | Republic of Cameroon | Pipeline |
| 6 | Republic of Chad | Pipeline |
| ... | ... | ... |
| 21540 | Africa | Dropped |
| 21541 | Africa | Closed |
| 21542 | Africa | Closed |
| 21543 | Africa | Closed |
| 21544 | Africa | Closed |

| | Last Stage Reached Name \ |
|-------|---------------------------|
| 2 | Bank Approved |
| 3 | Bank Approved |
| 4 | Concept Review |
| 5 | NaN |
| 6 | NaN |
| ... | ... |
| 21540 | NaN |
| 21541 | Bank Approved |
| 21542 | NaN |
| 21543 | Bank Approved |
| 21544 | Bank Approved |

| | Project Name \ |
|---|---|
| 2 | Rwanda Priority Skills for Growth (PSG) |
| 3 | Road Climate Resilience Project Second Additio... |
| 4 | Kenya Devolution and Service Delivery |

| | |
|-------|---|
| 5 | Cameroon Health System Strengthening Project |
| 6 | Chad Agribusiness and Rural Transformation Pro... |
| ... | ... |
| 21540 | UEMOA Regional Financial Sector Project |
| 21541 | Engineering and Technical Assistance Project |
| 21542 | Regional Development Project (03) |
| 21543 | REIMP(CEN.ENV.INFO) |
| 21544 | West Africa Pilot Community-based Natural Reso... |

| | |
|---|---|
| | Project Development Objective \ |
| 2 | The Program Development Objective is to expand... |
| 3 | The project will deliver sustainable climate r... |
| 4 | To improve citizen and government access to co... |
| 5 | To increase utilization and quality of reprodu... |
| 6 | The project Development Objective (PDO) is "to... |

| | |
|-------|-----|
| ... | ... |
| 21540 | NaN |
| 21541 | NaN |
| 21542 | NaN |
| 21543 | NaN |
| 21544 | NaN |

| | | | |
|-------|----------------------|------------------------|------------------|
| | Board Approval Date | Project Closing Date | Financing Type \ |
| 2 | 2017-07-05T00:00:00Z | 9/30/2023 12:00:00 AM | IDA |
| 3 | 2017-04-27T00:00:00Z | NaN | NaN |
| 4 | NaN | NaN | NaN |
| 5 | NaN | NaN | NaN |
| 6 | NaN | NaN | NaN |
| ... | ... | ... | ... |
| 21540 | NaN | NaN | NaN |
| 21541 | 1992-05-19T00:00:00Z | 12/31/1999 12:00:00 AM | IDA |
| 21542 | 1990-02-01T00:00:00Z | 12/31/1997 12:00:00 AM | NaN |
| 21543 | 1997-12-18T00:00:00Z | 6/30/2003 12:00:00 AM | Other |
| 21544 | 1995-09-14T00:00:00Z | 6/30/2004 12:00:00 AM | Other |

| | | | |
|-------|----------------------|-----------------|------------------|
| | Current Project Cost | IBRD Commitment | IDA Commitment \ |
| 2 | 120000000 | 0 | 120000000 |
| 3 | 39140000 | 0 | 35200000 |
| 4 | 340000 | 0 | 0 |
| 5 | 100000000 | 0 | 100000000 |
| 6 | 150000000 | 0 | 150000000 |
| ... | ... | ... | ... |
| 21540 | 10000000 | 0 | 10000000 |
| 21541 | 6100000 | 0 | 5500000 |
| 21542 | 55000000 | 15000000 | 40000000 |
| 21543 | 19700000 | 0 | 0 |
| 21544 | 7000000 | 0 | 0 |

| | |
|---|--|
| | Total IDA and IBRD Commitment Grant Amount \ |
| 2 | 120000000 0 |

| | | |
|-------|-----------|---------|
| 3 | 35200000 | 0 |
| 4 | 0 | 340000 |
| 5 | 100000000 | 0 |
| 6 | 150000000 | 0 |
| ... | ... | ... |
| 21540 | 10000000 | 0 |
| 21541 | 5500000 | 0 |
| 21542 | 55000000 | 0 |
| 21543 | 0 | 4100000 |
| 21544 | 0 | 4400000 |

| | Lending Instrument | Environmental Assessment Category |
|-------|-------------------------------|-----------------------------------|
| 2 | Program-for-Results Financing | NaN |
| 3 | Investment Project Financing | B |
| 4 | Investment Project Financing | C |
| 5 | Investment Project Financing | NaN |
| 6 | Investment Project Financing | NaN |
| ... | ... | ... |
| 21540 | Specific Investment Loan | C |
| 21541 | Technical Assistance Loan | C |
| 21542 | Financial Intermediary Loan | B |
| 21543 | Specific Investment Loan | C |
| 21544 | Specific Investment Loan | B |

| | Sector 1 | \ |
|-------|-----------------------------------|-------|
| 2 | Public Administration - Education | |
| 3 | Rural and Inter-Urban Roads | |
| 4 | | NaN |
| 5 | | NaN |
| 6 | | NaN |
| ... | | ... |
| 21540 | | NaN |
| 21541 | | Power |
| 21542 | Banking Institutions | |
| 21543 | Other Public Administration | |
| 21544 | Sub-National Government | |

```

                Sector 2 \
2                Tertiary Education
3                Public Administration - Transportation
4                NaN
5                NaN
6                NaN
...
21540            NaN
21541            NaN
21542            Other Agriculture, Fishing and Forestry
21543 Other Information and Communications Technologies
21544            Social Protection

```

```

                Theme 1
2                Education for the knowledge economy
3                NaN
4    Public expenditure, financial management and p...
5                NaN
6                NaN
...
21540            NaN
21541            Legal institutions for a market economy
21542            Environmental policies and institutions
21543            Land administration and management
21544            Rural services and infrastructure

```

[21543 rows x 18 columns]

df_new2.dtypes

```

Country                object
Project Status          object
Last Stage Reached Name object
Project Name            object
Project Development Objective object
Board Approval Date     object
Project Closing Date     object
Financing Type           object
Current Project Cost     object
IBRD Commitment          object
IDA Commitment           object
Total IDA and IBRD Commitment object
Grant Amount            object
Lending Instrument       object
Environmental Assessment Category object
Sector 1                object
Sector 2                object
Theme 1                 object
dtype: object

```

now i fill null value where the column which have integer data type

```
df_new2['Current Project Cost'].fillna(int(df_new2['Current Project Cost'].mean()), inplace=True)
```

```
df_new2['IDA Commitment'] = df_new2['IDA Commitment'].astype('float')
```

```
df_new2['Total IDA and IBRD Commitment'] = df_new2['Total IDA and IBRD Commitment'].astype('float')
```

```
df_new2['Grant Amount'] = df_new2['Grant Amount'].astype('float')
```

```
df_new2.head()
```

| | Country | Project Status | Last Stage |
|----------------|------------------------------------|----------------|------------|
| Reached Name \ | | | |
| 2 | Republic of Rwanda | Active | Bank |
| Approved | | | |
| 3 | Democratic Republic of Timor-Leste | Active | Bank |
| Approved | | | |
| 4 | Republic of Kenya | Dropped | Concept |
| Review | | | |
| 5 | Republic of Cameroon | Pipeline | |
| NaN | | | |
| 6 | Republic of Chad | Pipeline | |
| NaN | | | |

| | Project Name |
|---|---|
| 2 | Rwanda Priority Skills for Growth (PSG) |
| 3 | Road Climate Resilience Project Second Additio... |
| 4 | Kenya Devolution and Service Delivery |
| 5 | Cameroon Health System Strengthening Project |
| 6 | Chad Agribusiness and Rural Transformation Pro... |

| | Project Development Objective | Board Approval |
|--------|---|----------------------|
| Date \ | | |
| 2 | The Program Development Objective is to expand... | 2017-07-05T00:00:00Z |
| 3 | The project will deliver sustainable climate r... | 2017-04-27T00:00:00Z |
| 4 | To improve citizen and government access to co... | |
| NaN | | |
| 5 | To increase utilization and quality of reprodu... | |
| NaN | | |
| 6 | The project Development Objective (PDO) is "to... | |
| NaN | | |

| | Project Closing Date | Financing Type | Current Project Cost |
|---|-----------------------|----------------|----------------------|
| 2 | 9/30/2023 12:00:00 AM | IDA | 120000000 |
| 3 | NaN | NaN | 39140000 |
| 4 | NaN | NaN | 340000 |
| 5 | NaN | NaN | 100000000 |
| 6 | NaN | NaN | 150000000 |

| | IBRD Commitment | IDA Commitment | Total IDA and IBRD Commitment \ |
|---|-----------------|----------------|---------------------------------|
| 2 | 0 | 120000000.0 | 120000000.0 |
| 3 | 0 | 35200000.0 | 35200000.0 |
| 4 | 0 | 0.0 | 0.0 |
| 5 | 0 | 100000000.0 | 100000000.0 |
| 6 | 0 | 150000000.0 | 150000000.0 |

| | Grant Amount | Lending Instrument \ |
|---|--------------|-------------------------------|
| 2 | 0.0 | Program-for-Results Financing |
| 3 | 0.0 | Investment Project Financing |
| 4 | 340000.0 | Investment Project Financing |
| 5 | 0.0 | Investment Project Financing |
| 6 | 0.0 | Investment Project Financing |

| | Environmental Assessment Category | Sector 1 |
|---|-----------------------------------|-----------------------------------|
| 2 | NaN | Public Administration - Education |
| 3 | B | Rural and Inter-Urban Roads |
| 4 | C | NaN |
| 5 | NaN | NaN |
| 6 | NaN | NaN |

| | Sector 2 \ |
|---|--|
| 2 | Tertiary Education |
| 3 | Public Administration - Transportation |
| 4 | NaN |
| 5 | NaN |
| 6 | NaN |

| | Theme 1 |
|---|---|
| 2 | Education for the knowledge economy |
| 3 | NaN |
| 4 | Public expenditure, financial management and p... |
| 5 | NaN |
| 6 | NaN |

#fixing my column 'IBRD Commitment' keyword error

df_new2

| | Country | Project Status \ |
|---|------------------------------------|------------------|
| 2 | Republic of Rwanda | Active |
| 3 | Democratic Republic of Timor-Leste | Active |
| 4 | Republic of Kenya | Dropped |

| | | |
|-------|----------------------|----------|
| 5 | Republic of Cameroon | Pipeline |
| 6 | Republic of Chad | Pipeline |
| ... | ... | ... |
| 21540 | Africa | Dropped |
| 21541 | Africa | Closed |
| 21542 | Africa | Closed |
| 21543 | Africa | Closed |
| 21544 | Africa | Closed |

| | |
|-------|---------------------------|
| | Last Stage Reached Name \ |
| 2 | Bank Approved |
| 3 | Bank Approved |
| 4 | Concept Review |
| 5 | NaN |
| 6 | NaN |
| ... | ... |
| 21540 | NaN |
| 21541 | Bank Approved |
| 21542 | NaN |
| 21543 | Bank Approved |
| 21544 | Bank Approved |

| | |
|-------|---|
| | Project Name \ |
| 2 | Rwanda Priority Skills for Growth (PSG) |
| 3 | Road Climate Resilience Project Second Additio... |
| 4 | Kenya Devolution and Service Delivery |
| 5 | Cameroon Health System Strengthening Project |
| 6 | Chad Agribusiness and Rural Transformation Pro... |
| ... | ... |
| 21540 | UEMOA Regional Financial Sector Project |
| 21541 | Engineering and Technical Assistance Project |
| 21542 | Regional Development Project (03) |
| 21543 | REIMP(CEN.ENV.INFO) |
| 21544 | West Africa Pilot Community-based Natural Reso... |

| | |
|-------|---|
| | Project Development Objective \ |
| 2 | The Program Development Objective is to expand... |
| 3 | The project will deliver sustainable climate r... |
| 4 | To improve citizen and government access to co... |
| 5 | To increase utilization and quality of reprodu... |
| 6 | The project Development Objective (PDO) is "to... |
| ... | ... |
| 21540 | NaN |
| 21541 | NaN |
| 21542 | NaN |
| 21543 | NaN |
| 21544 | NaN |

| | | | |
|---|----------------------|-----------------------|------------------|
| | Board Approval Date | Project Closing Date | Financing Type \ |
| 2 | 2017-07-05T00:00:00Z | 9/30/2023 12:00:00 AM | IDA |

| | | | | |
|-------|----------------------|------------------------|-----|-------|
| 3 | 2017-04-27T00:00:00Z | | NaN | NaN |
| 4 | | NaN | NaN | NaN |
| 5 | | NaN | NaN | NaN |
| 6 | | NaN | NaN | NaN |
| ... | | ... | ... | ... |
| 21540 | | NaN | NaN | NaN |
| 21541 | 1992-05-19T00:00:00Z | 12/31/1999 12:00:00 AM | | IDA |
| 21542 | 1990-02-01T00:00:00Z | 12/31/1997 12:00:00 AM | | NaN |
| 21543 | 1997-12-18T00:00:00Z | 6/30/2003 12:00:00 AM | | Other |
| 21544 | 1995-09-14T00:00:00Z | 6/30/2004 12:00:00 AM | | Other |

| | Current Project Cost | IBRD Commitment | IDA Commitment | \ |
|-------|----------------------|-----------------|----------------|---|
| 2 | 120000000 | 0 | 120000000.0 | |
| 3 | 39140000 | 0 | 35200000.0 | |
| 4 | 340000 | 0 | 0.0 | |
| 5 | 100000000 | 0 | 100000000.0 | |
| 6 | 150000000 | 0 | 150000000.0 | |
| ... | ... | ... | ... | |
| 21540 | 10000000 | 0 | 10000000.0 | |
| 21541 | 6100000 | 0 | 5500000.0 | |
| 21542 | 55000000 | 15000000 | 40000000.0 | |
| 21543 | 19700000 | 0 | 0.0 | |
| 21544 | 7000000 | 0 | 0.0 | |

| | Total IDA and IBRD Commitment | Grant Amount | \ |
|-------|-------------------------------|--------------|---|
| 2 | 120000000.0 | 0.0 | |
| 3 | 35200000.0 | 0.0 | |
| 4 | 0.0 | 340000.0 | |
| 5 | 100000000.0 | 0.0 | |
| 6 | 150000000.0 | 0.0 | |
| ... | ... | ... | |
| 21540 | 10000000.0 | 0.0 | |
| 21541 | 5500000.0 | 0.0 | |
| 21542 | 55000000.0 | 0.0 | |
| 21543 | 0.0 | 4100000.0 | |
| 21544 | 0.0 | 4400000.0 | |

| | Lending Instrument | Environmental Assessment Category |
|---|-------------------------------|-----------------------------------|
| 2 | Program-for-Results Financing | NaN |
| 3 | Investment Project Financing | B |
| 4 | Investment Project Financing | C |
| 5 | Investment Project Financing | NaN |
| 6 | Investment Project Financing | NaN |

| | | |
|-------|-----------------------------|-----|
| ... | ... | ... |
| 21540 | Specific Investment Loan | C |
| 21541 | Technical Assistance Loan | C |
| 21542 | Financial Intermediary Loan | B |
| 21543 | Specific Investment Loan | C |
| 21544 | Specific Investment Loan | B |

| | |
|-------|-----------------------------------|
| | Sector 1 \ |
| 2 | Public Administration - Education |
| 3 | Rural and Inter-Urban Roads |
| 4 | NaN |
| 5 | NaN |
| 6 | NaN |
| ... | ... |
| 21540 | NaN |
| 21541 | Power |
| 21542 | Banking Institutions |
| 21543 | Other Public Administration |
| 21544 | Sub-National Government |

| | |
|-------|---|
| | Sector 2 \ |
| 2 | Tertiary Education |
| 3 | Public Administration - Transportation |
| 4 | NaN |
| 5 | NaN |
| 6 | NaN |
| ... | ... |
| 21540 | NaN |
| 21541 | NaN |
| 21542 | Other Agriculture, Fishing and Forestry |
| 21543 | Other Information and Communications Technologies |
| 21544 | Social Protection |

| | |
|-------|---|
| | Theme 1 |
| 2 | Education for the knowledge economy |
| 3 | NaN |
| 4 | Public expenditure, financial management and p... |
| 5 | NaN |
| 6 | NaN |
| ... | ... |
| 21540 | NaN |
| 21541 | Legal institutions for a market economy |
| 21542 | Environmental policies and institutions |

21543 Land administration and management
21544 Rural services and infrastructure

[21543 rows x 18 columns]

df_new2.loc[10]

Country
Republic of Türkiye
Project Status
Pipeline
Last Stage Reached Name
NaN
Project Name Modern land and geospatial data
systems for gr...
Project Development Objective To develop three-dimensional (3D)
city models ...
Board Approval Date
NaN
Project Closing Date
NaN
Financing Type
NaN
Current Project Cost
200000000
IBRD Commitment
200000000
IDA Commitment
0.0
Total IDA and IBRD Commitment
200000000.0
Grant Amount
0.0
Lending Instrument Investment
Project Financing
Environmental Assessment Category
NaN
Sector 1
NaN
Sector 2
NaN
Theme 1
NaN
Name: 10, dtype: object

df_new2["IBRD Commitment "] = df_new2[["IBRD Commitment
"]].astype(float)

df_new2.dtypes

```

Country                object
Project Status          object
Last Stage Reached Name object
Project Name            object
Project Development Objective object
Board Approval Date     object
Project Closing Date    object
Financing Type          object
Current Project Cost    int64
IBRD Commitment         float64
IDA Commitment          float64
Total IDA and IBRD Commitment float64
Grant Amount           float64
Lending Instrument      object
Environmental Assessment Category object
Sector 1                object
Sector 2                object
Theme 1                 object
dtype: object

```

```
df_new2["Project Status"].value_counts()
```

```

Closed      15783
Active      3088
Dropped     1725
Pipeline     942
Name: Project Status, dtype: int64

```

```

df_new2['IBRD Commitment '].fillna(int(df_new2['IBRD Commitment '].mean()), inplace=True)
df_new2['IDA Commitment'].fillna(int(df_new2['IDA Commitment'].mean()), inplace=True)
df_new2['Total IDA and IBRD Commitment'].fillna(int(df_new2['Total IDA and IBRD Commitment'].mean()), inplace=True)
df_new2['Grant Amount'].fillna(int(df_new2['Grant Amount'].mean()), inplace=True)

```

```
df_new2.isna().sum()
```

```

Country                3
Project Status          5
Last Stage Reached Name 6346
Project Name            2
Project Development Objective 9964
Board Approval Date     2758
Project Closing Date    4958
Financing Type          8506
Current Project Cost     0

```

```

IBRD Commitment          0
IDA Commitment           0
Total IDA and IBRD Commitment  0
Grant Amount             0
Lending Instrument       250
Environmental Assessment Category  8461
Sector 1                 1916
Sector 2                 10367
Theme 1                  9288
dtype: int64

```

```
df_new2.dropna(subset = ["Project Status"],inplace=True)
```

```
df_new2.head()
```

| | Country | Project Status | Last Stage |
|----------------|------------------------------------|----------------|------------|
| Reached Name \ | | | |
| 2 | Republic of Rwanda | Active | Bank |
| Approved | | | |
| 3 | Democratic Republic of Timor-Leste | Active | Bank |
| Approved | | | |
| 4 | Republic of Kenya | Dropped | Concept |
| Review | | | |
| 5 | Republic of Cameroon | Pipeline | |
| NaN | | | |
| 6 | Republic of Chad | Pipeline | |
| NaN | | | |

| | Project Name \ |
|---|---|
| 2 | Rwanda Priority Skills for Growth (PSG) |
| 3 | Road Climate Resilience Project Second Additio... |
| 4 | Kenya Devolution and Service Delivery |
| 5 | Cameroon Health System Strengthening Project |
| 6 | Chad Agribusiness and Rural Transformation Pro... |

| | Project Development Objective | Board Approval |
|--------|---|----------------------|
| Date \ | | |
| 2 | The Program Development Objective is to expand... | 2017-07-05T00:00:00Z |
| 3 | The project will deliver sustainable climate r... | 2017-04-27T00:00:00Z |
| 4 | To improve citizen and government access to co... | |
| NaN | | |
| 5 | To increase utilization and quality of reprodu... | |
| NaN | | |
| 6 | The project Development Objective (PDO) is "to... | |
| NaN | | |

| | Project Closing Date | Financing Type | Current Project Cost \ |
|---|-----------------------|----------------|------------------------|
| 2 | 9/30/2023 12:00:00 AM | IDA | 120000000 |

| | | | |
|---|-----|-----|-----------|
| 3 | NaN | NaN | 39140000 |
| 4 | NaN | NaN | 340000 |
| 5 | NaN | NaN | 100000000 |
| 6 | NaN | NaN | 150000000 |

| | IBRD Commitment | IDA Commitment | Total IDA and IBRD Commitment \ |
|---|-----------------|----------------|---------------------------------|
| 2 | 0.0 | 120000000.0 | 120000000.0 |
| 3 | 0.0 | 35200000.0 | 35200000.0 |
| 4 | 0.0 | 0.0 | 0.0 |
| 5 | 0.0 | 100000000.0 | 100000000.0 |
| 6 | 0.0 | 150000000.0 | 150000000.0 |

| | Grant Amount | Lending Instrument \ |
|---|--------------|-------------------------------|
| 2 | 0.0 | Program-for-Results Financing |
| 3 | 0.0 | Investment Project Financing |
| 4 | 340000.0 | Investment Project Financing |
| 5 | 0.0 | Investment Project Financing |
| 6 | 0.0 | Investment Project Financing |

| | Environmental Assessment Category | Sector 1 |
|---|-----------------------------------|-----------------------------------|
| 2 | NaN | Public Administration - Education |
| 3 | B | Rural and Inter-Urban Roads |
| 4 | C | NaN |
| 5 | NaN | NaN |
| 6 | NaN | NaN |

| | Sector 2 \ |
|---|--|
| 2 | Tertiary Education |
| 3 | Public Administration - Transportation |
| 4 | NaN |
| 5 | NaN |
| 6 | NaN |

| | Theme 1 |
|---|---|
| 2 | Education for the knowledge economy |
| 3 | NaN |
| 4 | Public expenditure, financial management and p... |
| 5 | NaN |
| 6 | NaN |

df_ex=pd.DataFrame()

df_ex[["Last Stage Reached Name","Financing Type","Current Project Cost","IBRD Commitment ","IDA Commitment","Total IDA and IBRD

```
Commitment","Grant Amount","Lending Instrument","Environmental
Assessment Category","Project Status"]]=df_new2[["Last Stage Reached
Name","Financing Type","Current Project Cost","IBRD Commitment ","IDA
Commitment","Total IDA and IBRD Commitment","Grant Amount","Lending
Instrument","Environmental Assessment Category","Project Status"]]
```

df_ex

| | Last Stage Reached Name | Financing Type | Current Project Cost \ |
|-------|-------------------------|----------------|------------------------|
| 2 | Bank Approved | IDA | 120000000 |
| 3 | Bank Approved | NaN | 39140000 |
| 4 | Concept Review | NaN | 340000 |
| 5 | NaN | NaN | 100000000 |
| 6 | NaN | NaN | 150000000 |
| ... | ... | ... | ... |
| 21540 | NaN | NaN | 10000000 |
| 21541 | Bank Approved | IDA | 6100000 |
| 21542 | NaN | NaN | 55000000 |
| 21543 | Bank Approved | Other | 19700000 |
| 21544 | Bank Approved | Other | 7000000 |

| | IBRD Commitment | IDA Commitment | Total IDA and IBRD Commitment |
|-------|-----------------|----------------|-------------------------------|
| \ | | | |
| 2 | 0.0 | 120000000.0 | 120000000.0 |
| 3 | 0.0 | 35200000.0 | 35200000.0 |
| 4 | 0.0 | 0.0 | 0.0 |
| 5 | 0.0 | 100000000.0 | 100000000.0 |
| 6 | 0.0 | 150000000.0 | 150000000.0 |
| ... | ... | ... | ... |
| 21540 | 0.0 | 10000000.0 | 10000000.0 |
| 21541 | 0.0 | 5500000.0 | 5500000.0 |
| 21542 | 15000000.0 | 40000000.0 | 55000000.0 |
| 21543 | 0.0 | 0.0 | 0.0 |
| 21544 | 0.0 | 0.0 | 0.0 |

| | Grant Amount | Lending Instrument \ |
|---|--------------|-------------------------------|
| 2 | 0.0 | Program-for-Results Financing |
| 3 | 0.0 | Investment Project Financing |

| | | |
|-------|-----------|------------------------------|
| 4 | 340000.0 | Investment Project Financing |
| 5 | 0.0 | Investment Project Financing |
| 6 | 0.0 | Investment Project Financing |
| ... | ... | ... |
| 21540 | 0.0 | Specific Investment Loan |
| 21541 | 0.0 | Technical Assistance Loan |
| 21542 | 0.0 | Financial Intermediary Loan |
| 21543 | 4100000.0 | Specific Investment Loan |
| 21544 | 4400000.0 | Specific Investment Loan |

| | Environmental Assessment Category | Project Status |
|-------|-----------------------------------|----------------|
| 2 | NaN | Active |
| 3 | B | Active |
| 4 | C | Dropped |
| 5 | NaN | Pipeline |
| 6 | NaN | Pipeline |
| ... | ... | ... |
| 21540 | C | Dropped |
| 21541 | C | Closed |
| 21542 | B | Closed |
| 21543 | C | Closed |
| 21544 | B | Closed |

[21538 rows x 10 columns]

```
df_ex["Project Status"].value_counts()
```

```
Closed      15783
Active      3088
Dropped     1725
Pipeline      942
Name: Project Status, dtype: int64
```

```
df_ex.isna().sum()
```

```
Last Stage Reached Name      6344
Financing Type                8501
Current Project Cost          0
IBRD Commitment               0
IDA Commitment                0
Total IDA and IBRD Commitment 0
Grant Amount                  0
Lending Instrument            248
Environmental Assessment Category 8459
Project Status                0
dtype: int64
```

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
```

```
df_ex['Project Status']= le.fit_transform(df_ex['Project Status'])
```

```
df_ex['Last Stage Reached Name'].value_counts()
```

```
Bank Approved      13494
Concept Review      1151
Begin Negotiation    223
Begin Appraisal      201
Decision Meeting     125
Name: Last Stage Reached Name, dtype: int64
```

```
df_ex['Last Stage Reached Name'].unique()
```

```
array(['Bank Approved', 'Concept Review', nan, 'Begin Negotiation',
       'Decision Meeting', 'Begin Appraisal'], dtype=object)
```

```
df_ex["Last Stage Reached Name"].fillna("Bank Approved", inplace =
True)
```

```
df_ex['Financing Type'].value_counts()
```

```
Other      4835
IDA         3541
Grants      2816
IBRD        1845
Name: Financing Type, dtype: int64
```

```
df_ex["Financing Type"].fillna("Other", inplace = True)
```

```
df_ex['Lending Instrument'].value_counts()
```

```
Specific Investment Loan      8019
Investment Project Financing  5497
Technical Assistance Loan     1653
Sector Investment and Maintenance Loan 1299
Development Policy Lending    1250
Financial Intermediary Loan    821
Emergency Recovery Loan        578
Structural Adjustment Loan     564
Adaptable Program Loan        518
Sector Adjustment Loan         387
Program-for-Results Financing  337
Learning and Innovation Loan   180
UNIDENTIFIED                   56
Programmatic Structural Adjustment Loan 51
Poverty Reduction Support Credit 40
Rehabilitation Loan           22
Debt and Debt Service Reduction Loan 16
Special Structural Adjustment Loan  2
Name: Lending Instrument, dtype: int64
```

```
mmm =[ 'Learning and Innovation Loan' , 'UNIDENTIFIED' , 'Programmatic
Structural Adjustment Loan' , 'Poverty Reduction Support Credit' ,
```

```
'Rehabilitation Loan' , 'Debt and Debt Service Reduction Loan' ,  
'Special Structural Adjustment Loan']
```

```
df_ex.loc[df_ex['Lending Instrument'].apply(lambda x:x in  
mmm),'Lending Instrument']= 'other_loan'
```

```
df_ex['Lending Instrument'].value_counts()
```

```
Specific Investment Loan      8019  
Investment Project Financing  5497  
Technical Assistance Loan     1653  
Sector Investment and Maintenance Loan  1299  
Development Policy Lending    1250  
Financial Intermediary Loan    821  
Emergency Recovery Loan       578  
Structural Adjustment Loan     564  
Adaptable Program Loan        518  
Sector Adjustment Loan        387  
other_loan                    367  
Program-for-Results Financing  337  
Name: Lending Instrument, dtype: int64
```

```
df_ex["Lending Instrument"].fillna("Other_loan", inplace = True)
```

```
df_ex['Environmental Assessment Category'].value_counts()
```

```
B      5859  
C      5210  
A      1156  
U       493  
F       333  
D        25  
H         2  
G         1  
Name: Environmental Assessment Category, dtype: int64
```

```
df_ex["Environmental Assessment Category"].fillna("B", inplace =  
True,)
```

```
df_ex.isna().sum()
```

```
Last Stage Reached Name      0  
Financing Type               0  
Current Project Cost         0  
IBRD Commitment              0  
IDA Commitment               0  
Total IDA and IBRD Commitment 0  
Grant Amount                 0  
Lending Instrument           0  
Environmental Assessment Category 0  
Project Status               0  
dtype: int64
```



```
df_ex.dtypes
```

```
Last Stage Reached Name      object
Financing Type                object
Current Project Cost          int64
IBRD Commitment               float64
IDA Commitment                float64
Total IDA and IBRD Commitment float64
Grant Amount                  float64
Lending Instrument            object
Environmental Assessment Category object
Project Status                int32
dtype: object
```

```
df_ex['Last Stage Reached Name']= le.fit_transform(df_ex['Last Stage Reached Name'])
```

```
df_ex['Financing Type']= le.fit_transform(df_ex['Financing Type'])
```

```
df_ex['Lending Instrument']= le.fit_transform(df_ex['Lending Instrument'])
```

```
df_ex['Environmental Assessment Category']= le.fit_transform(df_ex['Environmental Assessment Category'])
```

```
df_ex.head()
```

| | Last Stage Reached Name | Financing Type | Current Project Cost \ |
|---|-------------------------|----------------|------------------------|
| 2 | 0 | 2 | 120000000 |
| 3 | 0 | 3 | 39140000 |
| 4 | 3 | 3 | 340000 |
| 5 | 0 | 3 | 100000000 |
| 6 | 0 | 3 | 150000000 |

| | IBRD Commitment | IDA Commitment | Total IDA and IBRD Commitment \ |
|---|-----------------|----------------|---------------------------------|
| 2 | 0.0 | 120000000.0 | 120000000.0 |
| 3 | 0.0 | 35200000.0 | 35200000.0 |
| 4 | 0.0 | 0.0 | 0.0 |
| 5 | 0.0 | 100000000.0 | 100000000.0 |
| 6 | 0.0 | 150000000.0 | 150000000.0 |

| | Grant Amount | Lending Instrument | Environmental Assessment Category |
|---|--------------|--------------------|-----------------------------------|
| 2 | 0.0 | 6 | 1 |
| 3 | 0.0 | 4 | 1 |
| 4 | 340000.0 | 4 | 2 |
| 5 | 0.0 | 4 | 1 |
| 6 | 0.0 | 4 | 1 |

| | Project Status |
|---|----------------|
| 2 | 0 |
| 3 | 0 |
| 4 | 2 |
| 5 | 3 |
| 6 | 3 |

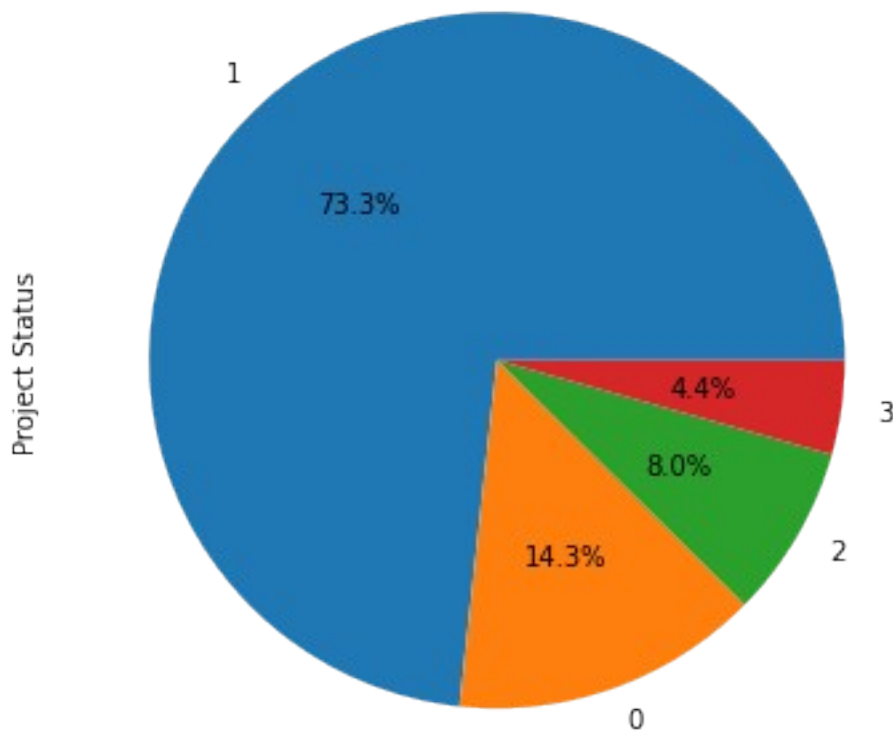
```
plt.figure(figsize=(30,20))
sns.pairplot(df_ex)
plt.show()
```

<Figure size 2160x1440 with 0 Axes>



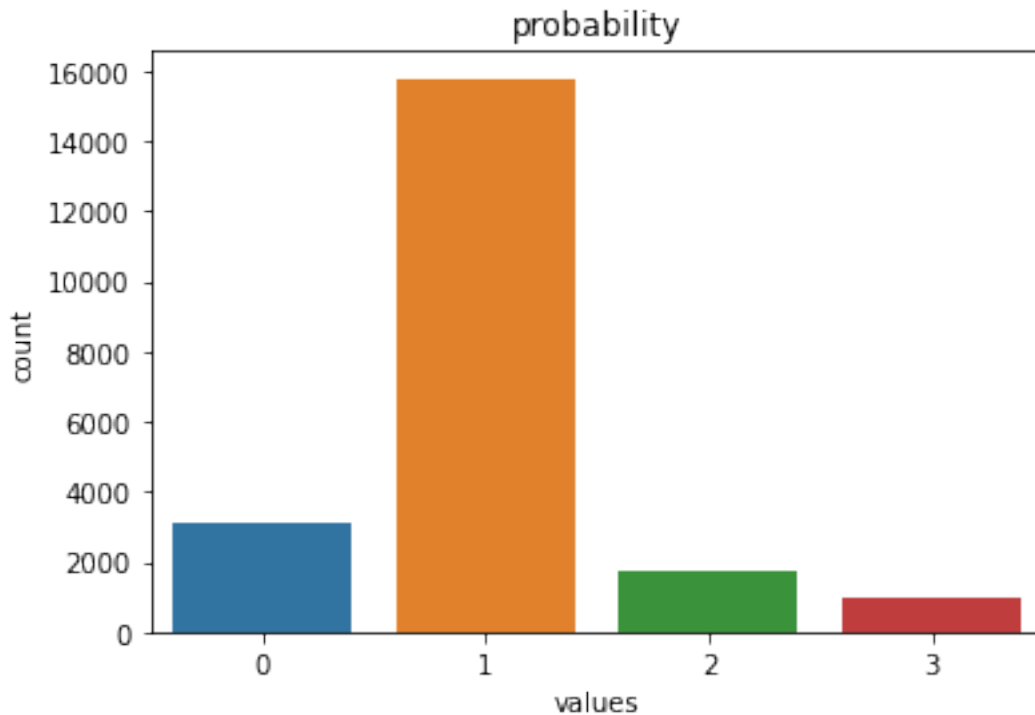
```
plt.figure(figsize=(9,6))
df_ex['Project Status'].value_counts().plot(kind='pie', autopct="%1.1f%")
```

<AxesSubplot:ylabel='Project Status'>



```
sns.countplot(df_ex['Project Status'])  
plt.title('probability')  
plt.xlabel('values')  
plt.ylabel('count')  
plt.show()
```

C:\Users\ACER\anaconda3\lib\site-packages\seaborn_decorators.py:36:
FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and
passing other arguments without an explicit keyword will result in an
error or misinterpretation.
warnings.warn(



```
df_ex["Project Status"].value_counts()
```

```
1    15783
0     3088
2     1725
3      942
```

```
Name: Project Status, dtype: int64
```

```
x=df_ex[['Last Stage Reached Name','Financing Type','Current Project Cost',
'IBRD Commitment ','IDA Commitment','Total IDA and IBRD Commitment',
'Grant Amount','Lending Instrument','Environmental Assessment Category']]
y=df_ex[['Project Status']]
```

```
x.shape,y.shape
```

```
((21538, 9), (21538, 1))
```

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.20,random_state=12)
```

```
x_train.shape,y_train.shape
```

```
((17230, 9), (17230, 1))
```

```
x_test.shape,y_test.shape
```

```
((4308, 9), (4308, 1))
```

```

from sklearn.tree import DecisionTreeClassifier
dt = DecisionTreeClassifier()

model_dt = dt.fit(x_train, y_train)

dt_pred = model_dt.predict(x_test)

dt_pred_proba = model_dt.predict_proba(x_test)
dt_pred_proba

array([[0., 1., 0., 0.],
       [0., 1., 0., 0.],
       [0., 1., 0., 0.],
       ...,
       [0., 0., 1., 0.],
       [0., 1., 0., 0.],
       [0., 1., 0., 0.]])

proba_df=pd.DataFrame(dt_pred_proba)

proba_df

   0    1    2    3
0  0.0  1.0  0.0  0.0
1  0.0  1.0  0.0  0.0
2  0.0  1.0  0.0  0.0
3  0.0  0.5  0.5  0.0
4  0.0  1.0  0.0  0.0
...
4303  0.0  0.0  1.0  0.0
4304  0.8  0.2  0.0  0.0
4305  0.0  0.0  1.0  0.0
4306  0.0  1.0  0.0  0.0
4307  0.0  1.0  0.0  0.0

[4308 rows x 4 columns]

proba_df[1]

0    1.0
1    1.0
2    1.0
3    0.5
4    1.0
...
4303  0.0
4304  0.2
4305  0.0
4306  1.0
4307  1.0
Name: 1, Length: 4308, dtype: float64

```

```

proba_df[1].value_counts()

1.000000    2732
0.000000     966
0.500000      61
0.666667      45
0.750000      33
...
0.869565        2
0.625000        2
0.882353        2
0.900000        2
0.866667        1
Name: 1, Length: 66, dtype: int64

from sklearn.metrics import
confusion_matrix,classification_report,accuracy_score

accuracy_score(dt_pred,y_test)

0.8391364902506964

print(confusion_matrix(dt_pred,y_test))

[[ 426  207    0   35]
 [ 169 2863   69   18]
 [    1   58  222   56]
 [   34   11   35  104]]

print(classification_report(dt_pred,y_test))

              precision    recall  f1-score   support

     0             0.68       0.64       0.66         668
     1             0.91       0.92       0.91        3119
     2             0.68       0.66       0.67         337
     3             0.49       0.57       0.52         184

 accuracy                   0.84         4308
 macro avg              0.69       0.69       0.69         4308
 weighted avg           0.84       0.84       0.84         4308

```

finding Z-score

```

from scipy.stats.mstats import zscore

df_ex['Project Status'].value_counts()

1    15783
0     3088
2     1725

```

```
3      942
Name: Project Status, dtype: int64

stats.zscore(df_ex['Project Status']).value_counts()

-0.038352    15783
-1.623803     3088
 1.547099     1725
 3.132551       942
Name: Project Status, dtype: int64
```